**Cloud computing (ITCS-6190) Project Deliverable 4**

**Group No: 9**

**Team members:**

Anudeep Billa

Aasish Chunduri

Govind Rahul Mathamsetti

Neethika Reddy Arepally

Lavanya Krishnan

**AWS Cost Analysis and Optimization**

Project Overview: This project involves analyzing correlations between music preferences and mental health metrics, with a focus on anxiety levels. AWS services played a crucial role in managing, processing, and analyzing the dataset, and in developing predictive models. The primary AWS services used were S3, AWS Glue, Athena, QuickSight, and SageMaker.

Cost Analysis:

1. Amazon S3:
   * Usage: S3 was used for storing raw data, processed data, and model artifacts.
   * Cost Drivers: Storage volume, number of read/write requests, and data transfer costs.
   * Analysis: Regular monitoring of stored data can help identify and delete outdated or unnecessary files to reduce storage costs.
2. AWS Glue:
   * Usage: Employed for ETL (Extract, Transform, Load) jobs to prepare data for analysis.
   * Cost Drivers: Number of Data Processing Units (DPUs) used and job run time.
   * Analysis: Optimizing ETL scripts and scheduling jobs efficiently can reduce runtime and DPU usage, thereby cutting costs.
3. Amazon Athena:
   * Usage: Utilized for querying datasets stored in S3.
   * Cost Drivers: Volume of data scanned per query.
   * Analysis: Effective query optimization, such as partitioning data and using compressed formats, can significantly reduce the amount of data scanned and thus lower costs.
4. Amazon QuickSight:
   * Usage: Used for visualizing data and gaining insights.
   * Cost Drivers: Number of SPICE (Super-fast, Parallel, In-memory Calculation Engine) capacity and user sessions.
   * Analysis: Regularly review SPICE capacity usage and optimize dashboards to use SPICE efficiently.
5. Amazon SageMaker:
   * Usage: Central to building, training, and deploying machine learning models.
   * Cost Drivers: Instance hours, data storage, and data processing.
   * Analysis: Optimize model training by selecting appropriate instance types, and manage resources effectively to minimize idle time.

Cost Optimization Strategies:

1. Right-Sizing and Reserved Instances: Analyze resource utilization and opt for right-sized instances to match workload requirements. Consider using Reserved Instances for services like SageMaker to save costs on predictable, long-term workloads.
2. Data Transfer Management: Minimize data transfer costs by optimizing data flow between services and reducing outbound data transfers.
3. AWS Cost Explorer:
   * Use AWS Cost Explorer to visualize and understand AWS spending patterns.
   * Set up custom reports to track the costs associated with each AWS service used in the project.
   * Leverage the Cost Explorer's recommendations for potential savings.
4. AWS Trusted Advisor:
   * Utilize AWS Trusted Advisor to get insights on cost optimization, performance, security, and fault tolerance.
   * Implement its recommendations on resource utilization and service limits.
5. Automating Resource Management: Implement scripts or use AWS services to automatically start and stop instances based on usage patterns, especially for SageMaker instances and ETL jobs in AWS Glue.
6. Delete Unnecessary Data: Regularly review and clean up S3 buckets and ETL job outputs to avoid paying for data storage that is no longer needed.
7. Use Amazon S3 Lifecycle Policies: Implement lifecycle policies to transition older data to cheaper storage classes or archive infrequently accessed data.

By following these strategies, you can optimize the costs of running AWS resources efficiently while maintaining the performance and scalability required for this data analysis project. Regular reviews and adjustments in line with changing project needs will ensure cost-effectiveness throughout the project lifecycle.